


Data filtering and genome assembly

 Reuben W Nowell  Timothy G Barraclough

Updated date: May 18, 2021

 An abbreviated version of this protocol was published in eLIFE in Feb 2021

Evolutionary dynamics of transposable elements in bdelloid rotifers

DOI: 10.7554/eLife.63194

Detailed protocol

Shell script for data filtering:

```
#!/bin/bash

echo 'Finding fastqs in '$1
echo 'Using '$2' CPUs and '${3}'G mem'

base=`basename $1`
outdir=${base}_bbduk
mkdir $outdir

zcat ${1}/*_1.fastq.gz > $outdir/${base}_1.raw.fq
zcat ${1}/*_2.fastq.gz > $outdir/${base}_2.raw.fq

pigz -p $2 $outdir/${base}*.raw.fq

echo "
ls -al $outdir/${base}_1.raw.fq.gz
ls -al $outdir/${base}_2.raw.fq.gz
echo "

bbduk.sh \
-Xmx${3}g \
threads=$2 \
in1=$outdir/${base}_1.raw.fq.gz \
in2=$outdir/${base}_2.raw.fq.gz \
out1=$outdir/${base}_1.bbduk.fq.gz \
out2=$outdir/${base}_2.bbduk.fq.gz \
ref=~/anaconda3/envs/main_env/opt/bbmap-38.22-0/resources/adapters.fa \
ktrim=r k=23 mink=11 hdist=2 maq=10 minlen=100 tpe tbo stats=${outdir}/bbduk.contaminants &> ${outdir}/bbduk.stats

echo "
kmercountexact.sh -Xmx${3}g threads=8 in1=$outdir/${base}_1.raw.fq.gz in2=$outdir/${base}_2.raw.fq.gz khist=$outdir/${base}.raw.khist

echo "
kmercountexact.sh -Xmx${3}g threads=8 in1=$outdir/${base}_1.bbduk.fq.gz in2=$outdir/${base}_2.bbduk.fq.gz
khist=$outdir/${base}.bbduk.khist

echo "
perl -lane 'if(/#/){next}else{print join(" ",$F[0],$F[1])}' $outdir/${base}.raw.khist > $outdir/${base}.raw.khist_jf.hist
perl -lane 'if(/#/){next}else{print join(" ",$F[0],$F[1])}' $outdir/${base}.bbduk.khist > $outdir/${base}.bbduk.khist_jf.hist
```

How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Nowell, R. and Barraclough, T. (2021). Data filtering and genome assembly. Bio-protocol Preprint. bio-protocol.org/prep1097.
2. Nowell, R. W., Wilson, C. G., Almeida, P., Schiffer, P. H., Fontaneto, D., Becks, L., Rodriguez, F., Arkhipova, I. R. and Barraclough, T. G. (2021). Evolutionary dynamics of transposable elements in bdelloid rotifers. eLIFE. DOI: [10.7554/eLife.63194](https://doi.org/10.7554/eLife.63194)

Copyright: Content may be subjected to copyright.